

Docket 100/114

-1-

Human antithrombin III, DNA sequences therefor, expression
vehicles and cloning vectors containing such sequences
and cell cultures transformed thereby, a process for
expressing human antithrombin III, and pharmaceutical
5 compositions comprising it.

Background of the Invention

Antithrombin III (ATIII) plays a critical role in
maintaining the fluidity of blood. Blood coagulation is
mediated by a series of serine proteases. Antithrombin III is
10 a potent inhibitor of Factors IXa, Xa (Kurachi et al.,
1976)¹, XI (Damus et al., 1973), XIIa (Stead et al., 1976);

15 ¹ Reference is made to the appended bibliography which is
hereby made a part hereof, the publications and other materials
there correspondingly cited in more detail being incorporated
herein by reference.

-2-

and thrombin (Rosenberg et al., 1973). Thus, ATIII regulates clot formation both at the activation level and the thrombin level. The physiological importance of ATIII in preventing excessive coagulation is revealed by studies of individuals whose antithrombin levels are decreased due to heredity (Egeberg, 1965; Odegard et al., 1977; Van der Meer et al., 1973; and Sas et al., 1974) or acquired deficiency (Abildgaard et al., 1970; Hannucci et al., 1973; Fagerhol et al., 1970). Such persons are prone to spontaneous thrombosis and the associated risks of disseminated intravascular coagulation (DIC), cardiac infarction, cerebrovascular accident, pulmonary embolism, etc. Transfusion of patients suffering from severe bleeding disorders complicated by DIC with antithrombin III concentrates obtained by blood fractionation has suggested that such replacement therapy can restore normal hemostatic function (Schipper et al., 1978).

Human antithrombin III is a single-chain glycoprotein synthesized by the liver and found in plasma at a concentration of approximately 20 mg/dl (Collen et al., 1977). Molecular weight determinations for antithrombin III have yielded values between 54,000 and 65,000 daltons (Rosenberg and Damus, 1973; Nordenman et al., 1977; Kurachi et al., 1976), of which some 10 percent is carbohydrate (Kurachi et al., 1976). The primary amino acid structure of human antithrombin III has been almost completely determined by Petersen et al. (1979) who reported that this protein has approximately 430 amino acid residues, 4 glucosamine-based oligosaccharide side units, and 3 disulfide bridges. However, these workers were uncertain as to the precise identity of several amino acids, and moreover, did not identify the sequence and length of an amino acid stretch in the middle of the protein due to incompletely overlapped peptides.

-3-

ATIII inactivates thrombin by forming a covalent 1:1 stoichiometric complex with the protease (Rosenberg and Damus, 1973; Owen, 1975). The anticoagulant effect of antithrombin III is enhanced by heparin, which greatly increases the rate of inhibitor-protease complex formation (Abildgaard, 1968).

It was thought possible that therapeutic administration of human antithrombin III produced by genetically engineered microorganisms would also be useful in the clinical prevention and management of thromboses. Such genetic engineering methods would provide sufficient quantities of material so as to enable its clinical testing as a demonstration of the safety and efficacy prerequisites to marketing. Therefore, the task was undertaken of cloning the gene for human antithrombin III and expressing it in a host cell.

Summary of the Invention

The present invention is directed to the means and methods of producing human antithrombin III via recombinant DNA technology, including 1) the discovery and identity of the entire DNA sequence of the mature protein as well as its signal polypeptide, and the 3'- and 5'-flanking regions thereof; 2) the construction of cloning and expression vehicles comprising said DNA sequence, enabling the amplification, and thence, expression of the mature human antithrombin III protein, as well as Met, fusion, or signal N-terminus conjugates thereof; and 3) viable cell cultures, genetically altered by virtue of their harboring such vehicles and capable of producing human antithrombin III polypeptide. Further, this invention provides human antithrombin III in physical state distinct from its existence in, or isolation from, a natural environment or source, it by virtue of its method of preparation herein, being essentially free of usual endogenous proteins and other native materials or substances.

-4-

This invention is directed to the recombinant DNA production of human antithrombin III in all of its aspects, and is not to be construed as limited to any specific details described herein and embraced within the compass of this invention. For example, the term "mature" as used herein connotes human ATIII as well as methionyl as a first amino acid, present by virtue of the ATG translational codon in the expression vector construction hereof.

10 Description of Preferred Embodiments

The work described herein was performed employing, inter alia, the microorganism E. coli K-12 strain 294 (end A, thi⁻, hsr⁻, _khsm⁺), as described in British Patent Application Publication No. 2055382A.

15 This strain has been deposited with the American Type Culture Collection, ATCC Accession No. 31446, on October 28, 1978. However, various other microbial strains are useful, including known E. coli strains such as E. coli B, E. coli x1776 (ATCC No. 31537, deposited July 3, 1979) and E. coli
20 W3110 (F⁻, λ⁻, protrophic) (ATCC No. 27325), or other microbial strains many of which are deposited and (potentially) available from recognized microorganism depository institutions, such as the American Type Culture Collection (ATCC)--cf. the ATCC catalogue listing. See German
25 Offenlegungsschrift 2644432. These other microorganisms include, for example, Bacilli such as Bacillus subtilis and other enterobacteriaceae among which can be mentioned as examples Salmonella typhimurium and Serratia marcesans, utilizing plasmids that can replicate and express heterologous
30 gene sequences therein.

Expression plasmids for bacterial use, e.g., E. coli are commonly derived using pBR322 as vector and appropriately

-5-

inserting the heterologous gene sequence together with translational start and stop signals in operable reading phase with a functional promoter, taking advantage of common or synthetically created restriction sites. The vector will carry
5 one or more phenotypic selection characteristic genes and an origin of replication to insure amplification within the host. Again, the heterologous insert can be aligned so as to be expressed together with a fused presequence, derivable for example from the trp system genes.

10 The present invention may also employ various yeast strains, hosting compatible expression vectors, such as the plasmid YRp7 (see Stinchcomb et al., Nature 282, 39 (1979)), which is capable of selection and replication in both E. coli and yeast, particularly Saccharomyces cerevisiae. A useful
15 strain is strain RH218 (Mioggari et al., J. Bacteriology 134, 48 (1978)) deposited with the American Type Culture Collection without restriction (ATCC No. 44076).

To express a heterologous gene such as the cDNA for human antithrombin III in yeast, it is necessary to construct a
20 plasmid vector containing four components. The first component is the part which allows for transformation of both E. coli and yeast and thus must contain a selectable gene from each organism. This can be the gene for ampicillin resistance from E. coli and the gene TRP1 from yeast. This component also
25 requires an origin of replication from both organisms to be maintained as a plasmid DNA in both organisms. This can be the E. coli origin from pBR322 and the ars1 origin from chromosome III of yeast or the origin of replication from 2 μ circle DNA.

The second component of the plasmid is a 5'-flanking
30 sequence from a highly expressed yeast gene to promote transcription of a downstream-placed structural gene. The

-6-

5'-flanking sequence can be that from the yeast 3-phosphoglycerate kinase (PGK) gene. The fragment is constructed in such a way so as to remove the ATG of the PGK structural sequence, replaced with a sequence containing alternative
5 restriction sites, such as XbaI and EcoRI restriction sites, for convenient attachment of this 5'-flanking sequence to the structural gene.

The third component of the system is a structural gene constructed in such a manner that it contains both an ATG
10 translational start and translational stop signals.

The fourth component is a yeast DNA sequence containing the 3'-flanking sequence of a yeast gene, which contains the proper signals for transcription termination and polyadenylation.

For example, plasmids directing the production of
15 methionyl-ATIII and pre ATIII in yeast can be constructed by respectively inserting gene fragments for the mature protein and mature protein plus signal peptide into the EcoRI site of 8.6 kbp expression plasmid YEplPT. (Hitzeman et al., Proc. of Berkeley Workshop on Recent Advances in Yeast Molecular
20 Biology, 20-22 May '82, University of California, Berkeley.) : The YEplPT vector contains a pBR322 origin to allow bacterial replication as well as a yeast 2 μ origin. YEplPT also carries an ampicillin resistance marker as a selectable trait in E. coli and the yeast TRP1 gene which is used to maintain the
25 plasmid in tryptophan auxotrophic yeast strains. A unique EcoRI cloning site is immediately downstream of a 1.6 kb yeast DNA fragment carrying the efficient promoter for the 3-phosphoglycerate kinase (PGK) gene. The ATIII structural gene fragments generated by complete EcoRI and partial PstI
30 digestion of pATIII-E7 or pATIII-J4 (see infra.) can be inserted into the YEplPT EcoRI site adjacent to the PGK

-7-

promoter. A 247 bp PstI-EcoRI DNA fragment derived from yeast 2 μ circle DNA is useful to convert the PstI sites at the 3' ends of the ATIII inserts into EcoRI ends; this convertor fragment also provides 2 μ terminator function.

5 Similarly, the present inventions can employ various cell culture systems with appropriate vectors. One useful host for the production of heterologous protein is the COS-7 line of monkey kidney fibroblasts (Gluzman, Cell 23, 175 (1981)).

However, the present invention could be practiced in any cell
10 line that is capable of the replication and expression of a compatible vector, e.g., WI38, BHK, 3T3, CHO, VERO, and HeLa cell lines. Additionally, what is required of the expression vector is an origin of replication and a promoter located in front of the gene to be expressed, along with any necessary
15 ribosome binding sites, RNA splice sites, polyadenylation site, and transcriptional terminator sequences. It will be understood that this invention, although described herein in terms of a preferred embodiment, should not be construed as limited to these sequences. For example, the origin of
20 replication of SV40 and other viral (e.g., Polyoma, Adeno, VSV, BPV, and so forth) vectors could be used, as well as cellular origins of DNA replication which could function in a nonintegrated state.

The strategy for the synthesis of heterologous peptide in
25 mammalian cell culture relies on the development of a vector capable of expression of a foreign gene under the control of a transcriptional unit. The replication of this vector in tissue culture can be accomplished by providing a DNA replication origin (such as from SV40 virus), and providing helper function
30 (T antigen) by the introduction of the vector into a cell line endogenously expressing this antigen (Lusky et al., Nature 293,

-8-

79 (1981)). A promoter of SV40 virus precedes the structural gene and ensures the transcription of the gene. Alternatively, the expressed gene can be incorporated into the cellular genome via procedures known per se so as to ensure stable transmission
5 of the gene.

A useful vector to obtain expression consists of pBR322 sequences which provides a selectable marker for selection in E. coli (ampicillin resistance) as well as an E. coli origin of DNA replication. These sequences are derivable from the
10 plasmid pML-1 (Lusky et al., Nature 293, 79 (1981)). The SV40 origin is derivable from a 342 base pair PvuII-HindIII fragment encompassing this region (Fiers et al., Nature 273, 113 (1978)) (both ends being convertible to EcoRI ends). These sequences, in addition to comprising the viral origin of DNA replication,
15 encode the promoter for both the early and late transcriptional unit. The orientation of the SV40 origin region is such that the promoter for the early transcriptional unit is positioned proximal to the gene encoding human antithrombin III.

20 Brief Description of the Drawings

Figure 1 depicts antithrombin III mRNA and cDNA clones. The top line represents the mRNA coding for human antithrombin III. DNAs employed as primers and hybridization probes are indicated by bars and the letters a-d above the mRNA. a) pool
25 of eight N-14 mers, b) pool of eight i-16 mers, c) i-14mer-
d) 230 bp Dde I fragment. The initial and final amino acid codons of the mature ATIII polypeptide are indicated by the circled 1 and 432 respectively. Key restriction endonuclease sites are shown by vertical lines. An approximate size scale
30 in nucleotides is included.

Figure 2 shows the nucleotide and amino acid sequence of human antithrombin III. The nucleotide sequence of the human

-9-

ATIII mRNA was determined from DNA sequence analysis of the cDNA clones pA62 and pA68. Predicted amino acids of the signal and mature ATIII polypeptide are shown above the DNA sequence and are numbered from the first residue of the mature protein.

5 The nucleotide sequence shown does not extend to the true 5' terminus of ATIII mRNA.

Figure 3 depicts the construction of pATIII-E7, details of which are provided infra. (A) Synthesis of 45 bp EcoRI-HindIII fragment by DNA Pol I extension of overlapping synthetic
10 oligonucleotides. (B) The synthetic fragment from (A), indicated by an open box, was ligated into the EcoRI-HindIII sites of pBR322, whose genes for ampicillin (amp) and tetracycline (tet) resistance are shown by dark boxes. A 40 bp EcoRI-SacI fragment was isolated from pR10 and ligated to a 590
15 bp fragment (shown by cross-hatched box) isolated from pA62. (C) EcoRI-SacI and SacI-PstI (dotted box) fragments were ligated into the trp expression vector fragment of LeIFA trp 103 (Gray et al., Nature 295, 503 (1982)) to yield pTA2. (D) pTA2 is a tetracycline resistant plasmid containing ATIII
20 structural gene sequences under control of the E. coli trp operon. To correct a two nucleotide deletion found in the structural gene, the EcoRI-SacII fragment was replaced with fragments synthesized by different means. The resulting constructions, pATIII-E7 and pATIII-J4, resulted respectively
25 in synthesis of the human mature (with accommodation of methionyl first amino acid) ATIII and human pre ATIII polypeptides in E. coli cells.

Detailed Description

30 Methods:

RNA preparation. Human liver messenger RNA was prepared by the guanidinium thiocyanate method (Ullrich et al., 1977)

-10-

followed by oligo dT cellulose chromatography (poly(A)⁺ RNA), or by phenol extraction of polysomes isolated in the presence of the potent ribonuclease inhibitor, hydroxystilbamidine isethionate (Lizardi, 1980) (polysomal RNA). To obtain
5 approximately 3 mg of polysomal RNA, 5 g of liver were homogenized in 10 mM NaCl, 1.5 mM MgCl₂, 1 percent NP40, 250 mM sucrose, 10 mM Tris-HCl, pH 7.5, 1.5 mM hydroxystilbamidine isethionate (Merrell, Cincinnati, OH). Nuclei and membranes were cleared from the homogenate (750 xg, 5 m), and the
10 polysomes were spun thru 0.5 M sucrose in 0.3 M NaCl, 50 mM Tris-HCl, pH 7.5, 5 mM MgCl₂, 1.5 mM hydroxystilbamidine isethionate onto a pad of 1.0 M sucrose in the same buffer (12,000 xg, 20 m). The polysome band was phenol extracted and ethanol precipitated, resuspended in water and clarified by
15 centrifugation at 12,800 xg for 30 m. The resultant polysomal RNA preparations had A260/A280 ratios in the range 1.8 - 2.0.

Poly A(+) liver RNA was enriched for ATIII messages by centrifugation through a linear sucrose density gradient (15 percent - 30 percent sucrose in 0.1 M NaCl, 12.5 mM EDTA, 10 mM
20 Tris, pH 7.5, 0.15 percent Sarkosyl, Beckman SW50.1, 47 Krpm, 4 h, 20°C). Aliquots of the fractionated RNA were used to direct in vitro protein synthesis in a ³⁵S-methionine supplemented rabbit reticulocyte lysate (BRL, Bethesda, MD). ATIII mRNA-containing fractions were identified by immunoprecipitation
25 of in vitro synthesized protein with rabbit antihuman ATIII antiserum (Atlantic Antibodies, New Brunswick, Maine) (Kessler, 1976) and subsequent SDS gel electrophoresis (Laemmli, 1970).

Synthesis, Cloning and Screening of cDNA. Reverse transcription reactions (containing from 0.1 - 5 µg RNA) were
30 primed with the synthetic DNA fragments indicated in the text (approximately 200 ng) or oligo dT₍₁₂₋₁₈₎ (Collaborative)

-11-

(approximately 1 µg). Double stranded cDNA was prepared, oligo dC-tailed, inserted into PstI cleaved G-tailed pBR322 and transformed into E. coli 294 as previously described (Lawn et al., 1981).

5 E. coli transformants were screened with ³²P
5'-end-labeled synthetic oligonucleotides or ³²P-labeled DNA
prepared by random priming of given restriction fragments as
previously described in Lawn et al. (1981). The synthetic DNA
fragments used as primers and probes were prepared by the
10 phosphotriester method (Crea and Horn, 1980). Procedures for
DNA preparation and restriction enzyme analysis have also been
published (Lawn et al., 1981).

Hybridization Selection of ATIII mRNA. Approximately 4 µg
of a 230 bp DNA fragment from plasmid pA3 was coupled to 25 mg
15 DBM-cellulose using the procedure developed by Hoyer and Stark
(1975). This DNA cellulose was used to isolate ATIII-specific
mRNA from 1 mg of human liver polysomal RNA utilizing the
conditions described in Bock et al. (1982).

Construction of the expression plasmid. Procedures for DNA
20 fragment isolation and conditions for the ligation reactions
have been published elsewhere (Lawn et al., 1981) and are
applicable herein. Synthesis of fragments connecting the
promoter and 5' end of the gene are described below.

Synthesis of Eco RI - Hind III fragment. The principles
25 underlying the design of the 36-base (dCTAGAATTCTATGCACGGCTCG
CCAGTGGACATCTG) and 37-base (dCGCAAGCTTCCGCGGCTTGGCTGTGCAGAT
GTCCAAT) oligonucleotides used to synthesize the 45 bp EcoRI -
HindIII fragment are presented in detail infra. The 36-mer and
37-mer were phosphorylated at their 5' ends with T4
30 polynucleotide kinase (P-L), and 2 µg of each oligonucleotide
were annealed together for 10 min at 68°C, 12 min at 20°C and

-12-

12 m at 4°C in a volume of 40 µl. The annealing reaction was supplemented with deoxynucleotide triphosphates to 400 µM each and 5 units of E. coli DNA polymerase Klenow fragment (BRL) (total volume = 50 µl). The polymerization reaction was
5 incubated for 30 min at 20°C and 1 h at 37°C. Buffer for the polymerization reaction contained 50 mM NaCl, 6 mM Tris.HCl, pH 7.5, 6 mM Mg Cl₂, 5 mM DTT, and 100 µg/ml BSA; the annealing reaction contained these same components at 1.25 X strength. After polymerization, the reaction mixture was deproteinized by
10 phenol and chloroform extraction and digested sequentially with HindIII and EcoRI. The resultant 45 bp fragment was purified by electrophoresis in a 15 percent polyacrylamide gel.

Synthesis of 40 bp EcoRI - SacII fragment. A 1400 bp PstI fragment containing the 5' end of the ATIII structural gene was
15 isolated from pA62. 25 µg of this fragment and 3 µg of 5' phosphorylated 36-mer (used previously for synthesis of the EcoRI-HindIII fragment) were incubated in 33 µl H₂O at 100°C for 5 m, then quick frozen. This annealing reaction was thawed and adjusted to a final volume of 50 µl containing 20 mM KCl, 8
20 mM MgCl₂, 30 mM DTT, 20 mM Tris.HCl, pH 8.3, 500 µM each dGTP, dATP, dCTP and dTTP, and 25 units reverse transcriptase (BRL). The reaction was incubated for 10 m at 20°C and 50 m at 42°C. After deproteinization, the DNA was treated with 7 units E. coli DNA polymerase Klenow fragment (Boehringer) as
25 described previously. Digestion with SacII and EcoRI followed Klenow treatment and produced the desired 40 bp fragment which was gel isolated.

DNA sequencing. DNA sequence analysis was performed by the methods of Maxam and Gilbert (1980) and Sanger, Nicklen and
30 Coulson (1977).

Protein Analysis. Cell pellets containing 1 A₅₅₀ unit were obtained at various stages during the fermentation of

-13-

antithrombin expression plasmids ATIII-E7 and pATIII-J4 (transformed into E. coli strain W3110). Each sample was resuspended in buffer containing 40mM imidazole Cl, pH 7.5, 2 percent SDS, 10 percent glycerol and 5 percent BME, and boiled
5 for 3m. Samples were subjected to electrophoresis in the gels systems of Laemmli (1970) or Weber and Osborn (1969). Total protein was visualized by Coomassie Brilliant Blue staining and bacterially synthesized ATIII was identified by Western blot analysis (Renart et al., 1975).

10

Results:

Initial cDNA clone for human antithrombin III. A recombinant plasmid containing 600 bp of the ATIII structural gene was identified in a cDNA bank prepared from size
15 fractionated RNA by hybridization to synthetic DNA fragment probes.

The published partial protein sequence of human antithrombin III (Petersen et al., 1979) was inspected for regions with limited codon degeneracy. Two such regions were
20 identified and the corresponding synthetic DNAs of 14 or 16 nucleotides in length were prepared (Table I).

25

30

-14-

TABLE I

Amino Acid Position*	Protein Sequence	Synthetic DNA Sequence	Use
5 17	MetAsnProMetCys	5' dCACATAGGGTTCAT 3' G A T C	hybridization probe
243	MetMetTyrGlnGluGly	5' dCCCTCCTGGTACATCA 3' T T A	hybridization probe; cDNA primer
10	Synthetic DNA fragments used for cloning human antithrombin III		

*Numbering system of Petersen *et al.*, 1979.

The more C-terminally located of these sequences, a pool of
 15 eight 16 mers, was used to prime cDNA synthesis from poly(A)+
 RNA which had been enriched for ATIII mRNA. Previous *in vitro*
 translation studies have suggested that ATIII mRNA constitutes
 only 0.5 percent of bovine liver poly(A)+ RNA (MacGillivray
et al., 1979). We therefore enriched for ATIII encoding RNA by
 20 sucrose density gradient centrifugation, and identified the
 desired fraction by immunoprecipitation of ATIII from *in vitro*
 translation assays. Approximately 250 transformants were
 obtained upon insertion of the oligo dC-tailed, double stranded
 cDNA produced by specifically primed reverse transcription of
 25 the size fractionated RNA into *Pst*I cleaved, oligo dG-tailed
 pBR322. Colony hybridization of this clone bank to the ³²P
 5'-end-labeled 14- and 16-base synthetic nucleotides (Table I)
 revealed one transformant with strong hybridization to the
 16mer, but not to the 14mer. DNA from this plasmid, designated
 30 pA3, was subjected to sequence analysis, and its nucleotide

-15-

sequence was shown to correspond to the protein sequence of human antithrombin III extending from amino acid 239 towards the C terminus for approximately 200 residues (Fig. 1).

The identification of a clone encoding a portion of ATIII
5 3' to amino acid residue 243 was unexpected since the cDNA from which this clone bank was constructed had been primed with a pool of 16-base oligonucleotides predicted to encode the region around amino acid 243. Transcripts from the complementary primer were therefore expected to extend in the 5' direction
10 from amino acid 243 and should have not included sequences 3' to the probe. In actual fact, pA3 begins near to the C-terminus of the mature protein and extends for only 12 nucleotides 5' to the probe sequence, thus hybridizing with the 16mer pools, but not the 14mers (which are located in a more
15 N-terminal region of the protein). This particular cDNA clone perhaps resulted from the self-priming of ATIII RNA species for the reverse transcription reaction.

cDNA Clones for the Complete ATIII Structural Gene. An internal restriction fragment from pA3 was used to purify the
20 antithrombin III messenger RNA subsequently used for generating an oligo-dT primed cDNA clone bank. Two overlapping cDNA clones which together encode the entire human ATIII structural gene and its associated 5' and 3' untranslated regions were isolated from this bank.

25 DNA cellulose was prepared from the 230 bp DdeI fragment of pA3 and used to select complementary sequences from 1 mg of human polysomal liver RNA as outlined supra. In vitro translated protein from an aliquot of the bound RNA was considerably enriched for antithrombin III. The hybrid-
30 selected RNA was used to direct oligo-dT primed cDNA synthesis. Double stranded cDNA was again cloned into the PstI site of pBR322 through homopolymeric dGC tailing, this time

-16-

yielding about 500 transformants. Quadruplicate filters of the clone bank were made and screened with 4 probes distributed along the length of the ATIII structural gene (Fig. 1). The four hybridization probes were: (a) the pool of eight
5 N-terminal 14 mers (N-14 mers) synthesized for the initial screening experiment, (b) the pool of eight 16 mers also used in the initial screening (i-16 mers), (c) a 14-base synthetic DNA (i-14mer, dTGAGGACCATGGTG) whose sequence had been determined by sequencing of pA3 and was complementary to the
10 coding strand at amino acid 271-276 (numbering scheme as per Petersen *et al.*, 1979), and (d) the 230 bp *DdeI* fragment from pA3 (which had also been used for mRNA hybridization selection).

28 of the 500 colonies generated by oligo dT priming of
15 hybridization-selected RNA were positive when screened with the *DdeI* fragment used for RNA enrichment. The length of ATIII cDNAs inserted in these transformants was assessed by hybridization to the three more N-terminal probes: 20 were positive with the i-14 mer probe, 17 with the pool of 16 mers,
20 and 3 with the N-14 mers. The latter 3 plasmids had inserts of about 1.8 kb and were analyzed further. Partial DNA sequencing of these plasmids verified that they did indeed encode the amino terminus of mature human antithrombin III.

Subsequent restriction mapping and DNA sequencing of one
25 plasmid, pA62, revealed the primary structure of an ATIII mRNA composed of an extended 5' untranslated region and a coding region of 1390 bps. pA62, however, was incomplete at the 3' end of the gene. The homopolymeric dGC tails began after the second nucleotide of the codon for the carboxyl terminal amino
30 acid residue. Thus, this plasmid was missing the last nucleotide of the C-terminal codon, the stop codon and the 3' untranslated region.

-17-

Further analysis of partial length ATIII cDNA clones in the library constructed from oligo dT primed, hybridization-enriched mRNA revealed a plasmid, pA68, which contained 400 bps from the 3' end of the antithrombin III structural gene, an 84 bp 3' untranslated region, and a poly(A) tail. Sequence analysis of pA68 and the overlapping clone, pA62, has provided the complete primary structure of the human antithrombin III gene and portions of its flanking, untranslated regions. In addition, common restriction sites within the overlapping regions of the partial cDNA clones have been utilized to construct a full length structural gene which has been inserted into an expression plasmid so as to direct the synthesis of human antithrombin III under trp operator-promoter control in *E. coli*.

Sequence and Structure of the human antithrombin III messenger RNA. Nucleotide sequence analysis of the A62 and A68 cDNA clones reveals that ATIII mRNA has a 96 nucleotide signal sequence (32 amino acids), a 1296 nucleotide sequence encoding the mature protein (432 amino acids), and 84 residues in the 3' untranslated region.

DNA sequencing reveals that a signal peptide of 32 residues probably precedes the mature N terminus of human antithrombin III. This sequence displays two features commonly attributed to signal peptide regions (Jackson and Blobel, 1980). First, a very hydrophobic region (LeuLeuSerLeuLeuLeuIle) is observed about ten residues prior to the cleavage site. Second, the presence of a cysteine immediately prior to the mature N terminus is in agreement with the precedent that the amino acid residue immediately preceding the cleavage site has in all cases been one of the smallest amino acids.

The mature form of human antithrombin III contains 432 amino acid residues. The protein sequence determined by DNA

-18-

analysis agrees perfectly with the partial sequence published by Petersen et al. (1979) and has provided definitive sequence information for the regions where the original amino acid determinations were ambivalent or unknown. Amino acid residues 5 100, 101, 326 and 328 (numbering scheme as per Petersen) which were identified only as Glx by protein sequencing have been unambiguously determined as gln, gln, gln, and gln. Residue number 329 is an asp. DNA sequencing has also revealed that eight additional amino acids (ValLeuValAsnThrIleTyrPhe) occur 10 between Leu 213 and Lys 214 of the sequence, which Petersen et al. (1979) did not appreciate.

Finally, the nucleotide sequence of ATIII cDNA indicates that the carboxy terminus of this protein is not processed, since the C-terminal residue of the mature protein, lysine, is 15 immediately followed by a stop codon. The 3' noncoding region consists of 84 nucleotides following the termination codon UAA. The sequence AATAAA is present at minus 30 residues from the polyadenylation site.

20 Bacterial Expression of Human Antithrombin III

pATIII-E7, a 7 kb plasmid producing 50,000 dalton methionyl-ATIII, was constructed by placing segments of partial cDNA clones pA62 and pA68 behind the E. coli trp promotor (Fig. 3). Construction of this expression plasmid necessitated the 25 synthesis of a DNA fragment incorporating an EcoRI cleavage site, an ATG translation initiation codon and the nucleotide sequence encoding the amino terminus of the mature structural gene, through the first unique restriction site, SacII. We attempted to synthesize such a fragment by a combination of 30 chemical and enzymatic procedures. A 36 base deoxyoligo-nucleotide was synthesized which contained 3 arbitrarily chosen nucleotides followed by the six base EcoRI recognition .

-19-

sequence, then TATG and the 23 nucleotides encoding the first eight codons of mature human antithrombin III.

This synthetic 36-mer was designed to provide optimal spacing between the ribosome binding sequence and the start of translation, as well as to minimize secondary structure in the 5' end of the gene. To accomplish this, the triplets selected to encode amino acid residues 2, 3 and 4 do not correspond to the natural codons observed in the pA62 cDNA clone. A 37 base deoxyoligonucleotide which anneals to the 3' end of the 36 mer by 12 base pairs and extends towards the 3' terminus of the gene was also synthesized. Reading from its 3' to 5' end, the 37 mer contains nucleotides coding for amino acid residues 4 through 13 (containing a SacII recognition sequence) and an artificial HindIII site to be employed in propagating the linker fragment. The 36- and 37 mers were annealed, and the initial 12 bp long double stranded region was extended with DNA polymerase Klenow fragment. The duplex product was trimmed with EcoRI and HindIII and the purified EcoRI to HindIII fragment was inserted into pBR322. DNA sequence analysis of the insert from this construction, pR10, revealed that the synthetic DNA was missing 2 adjacent nucleotides of the desired sequence.

The deletion may result from improper extension of this particular set of synthetic DNAs by the DNA polymerase employed. A method for correcting the 2-base deletion is discussed below. Before this was developed, the fragment containing the deletion was utilized as a convenient EcoRI to SacII linker for constructing a direct expression plasmid.

The EcoRI to SacII fragment from pR10 was used to join segments of the pA62 and pA68 partial cDNA clones to the trp promoter in pBR322. Approximately 120 ng of the EcoRI to SacII fragment from pR10 was ligated to about 1 µg of the 590 bp

-20-

SacII to SacI fragment from pA62. A 630 bp EcoRI to SacI DNA fragment was gel isolated following restriction of the ligation mixture with the appropriate enzymes. This EcoRI to SacI fragment was joined to the 785 bp DNA fragment produced by SacI complete and PstI partial digestion of pA68, and a pBR322 expression vector having EcoRI and PstI ends in a three part ligation reaction. The expression vector contains a trp promoter-operator fragment which is inserted at the EcoRI site of pBR322 and reads towards the amp gene. An EcoRI sequence follows the promoter ribosome binding site; and the PstI site in the amp gene is used to fuse the vector to the 3' end of the structural gene. The product of this 3-part ligation is designated pTA2.

A derivative pTA2 was made by substituting its 38 bp EcoRI to SacII fragment with a 40 bp piece of synthetic DNA which contains an EcoRI site, initiation codon and intact coding sequence for amino acid residues 1 through 12 of ATIII. This 40 bp DNA fragment was synthesized by a process which utilized two different DNA polymerase enzymes. First, a 36 base long primer was hybridized to a template DNA fragment from pA62 and extended using the polymerase activity of AMV reverse transcriptase. (See supra. for a description of 36-mer.) Following primer elongation, the dual 3' to 5' exonuclease and 5' to 3' polymerase activities of E. coli Klenow fragment were exploited to trim off excess single stranded template DNA and polymerize the first several codons of the structural gene, the translation initiation codon and the EcoRI tail. The preparation was subsequently cut with EcoRI and SacII and the 40 bp fragment was inserted into pTA2. pATIII-E7, the resultant plasmid, was verified to have the correct coding information between the EcoRI and SacII sites by DNA sequence analysis.

-21-

Furthermore, pATIII-E7 produces a 50,000 dalton protein under trp control which has been immunologically identified as antithrombin III.

Additionally, pATIII-J4, a plasmid expressing
5 pre-antithrombin III under trp operator-promoter control has been constructed. pATIII-J4 was derived from pTA2 by replacement of the 38 bp Eco RI - Sac II fragment with a 145 bp Eco RI - Sac II fragment encoding the presumptive 36 residue signal sequence of antithrombin III and the first 12 aminoacids
10 of the mature polypeptide. This fragment was synthesized from a 24-base synthetic DNA primer (dCTAGAATTCATGTATTCCAATGTG) and the 1400 bp pA62 Pst I fragment by sequential DNA polymerase I, Sac II and Eco RI treatments (Lawn et al., 1981). pATIII-J4 expresses two proteins under trp control which react with ATIII
15 antisera. The major component has an apparent molecular weight approximating the calculated molecular weight of pre-antithrombin III. A minor band was present that comigrates with the ATIII band produced from pATIII-E7.

20 Inactivation of human thrombin by bacterially synthesized antithrombin III.

ATIII inactivates thrombin by reacting with it to form a covalent, 1:1 stoichiometric complex (Jesty, 1979). For example, cell extract of E. coli W3110 transformed with
25 pATIII-J4 can be incubated with human thrombin in the presence of heparin. Following polyacrylamide gel electrophoresis, ATIII thrombin complexes are detected by Western blotting (Renart, et al., 1979) with rabbit anti-human ATIII antiserum. The bacterial ATIII-human thrombin complex is observed in
30 reactions to which thrombin has been added.

Pharmaceutical Compositions

The compounds of the present invention can be formulated according to known methods to prepare pharmaceutically useful compositions, whereby the human antithrombin III product hereof
5 is combined in admixture with a pharmaceutically acceptable carrier vehicle. Suitable vehicles and their formulation, inclusive of other human proteins, e.g. human serum albumin or plasma preparations, are described for example in Remington's
Pharmaceutical Sciences Mark Publishing Co., Easton, Pa., 15th
10 2nd., 1975, which is hereby incorporated by reference. Such compositions will contain an effective amount of the protein hereof together with a suitable amount of vehicle in order to prepare pharmaceutically acceptable compositions suitable for effective administration, preferably parenteral, to the host.
15 Notwithstanding that reference has been made to particular preferred embodiments, it will be further understood that the present invention is not to be construed as limited to such, rather to the lawful scope of the appended claims.

Bibliography

1. Abildgaard, U. Scand. J. Haemat. 5: 440-453 (1968).
2. Abildgaard, U. et al. Scand. J. Clin. Lab. Invest. 26 (4): 349-354 (1970).
3. Bock, S.C. et al. Proc. Natl. Acad. Sci. USA 79: 1032-1036 (1982).
4. Collen, D. et al. Eur. J. Clin. Invest. 7: 27-35 (1977).
5. Crea, R. and Horn, T. Nucl. Acids Res. 8: 2331-2348 (1980).
6. Damus, P.S. et al. Nature 246: 355-357 (1973).
7. Egeberg, O. Thromb. Diath. Haemorrh. 13: 516-530 (1965).
8. Fagerhol, M.K. et al. Scand. J. Haematol 7 (1): 10-17 (1970).
9. Jackson, R.C. and Blobel, G. Annals N.Y. Acad. Sci. vol. 343 (1980).
10. Jesty, J. J. Biol. Chem. 254: 1044-1049 (1979).
11. Kessler, S.W. J. Immunology 117: 1482-1490 (1976).
12. Kurachi, K. et al. Biochemistry 15 (2): 373-377 (1976).
13. Laemmli, U.K. Nature 277: 680-685 (1970).
14. Lawn, R.M. et al. Nucl. Acids Res. 9: 6103-6114 (1981).
15. Lizardi, P.M. J. Cell. Biol. 87: 292-296 (1980).
16. Mannucci, L. et al. Scand. J. Gastroenterol. Suppl. 8 (19): 103-107 (1973).
17. Maxam, A.M. and Gilbert, W. Methods Enzymol. 65: 499-560 (1980).
18. Nordenman, B. et al. Eur. J. Biochem. 78 (1): 195-204 (1977).
19. Hoyer, B.E. and Stark, G.R. Cell 5: 301-310 (1975).
20. Odegard, O.R. et al. Scand. J. Haematol. 18: 86-90 (1977).
21. Owen, W.G. Biochimica et Biophysica Acta. 405: 380-387 (1975).
22. Petersen, T.E. et al. In The Physiological Inhibitors of Coagulation and Fibrinolysis. (D. Collen, B. Wiman and M. Verstraete, eds.) Elsevier, Amsterdam. p. 48 (1979).
23. Renart, J. et al. Proc. Natl. Acad. Sci. USA. 76: 3116-3120. (1979)
24. Rosenberg, R.D. et al. J. Biol. Chem. 248 (18): 6490-6505 (1973).

-24-

25. Sanger, F., Nicklen, S. and Coulson, A.R. Proc. Natl. Acad. Sci. USA 74: 5463-5467.
26. Sas, G. et al. Thrombos. Diathes. Haemorrh. (Stuttg.) 32: 105 (1979).
27. Schipper, H.G. et al. Lancet 1 (8069): 854-856 (1978).
28. Stead, N. et al. J. Biol. Chem. 251: 6481-6488.
29. Ullrich, A. et al. Science 196, 1313-1315 (1977).
30. Van der Meer, J. et al. J. Clin. Path. 26: 532-538 (1973).
31. Weber, K., and Osborn, M. J. Biol. Chem. 244: 4406-4412 (1969).

Claims:

1. Human antithrombin III as a product of genetically altered cell culture.
- 5 2. Human antithrombin III essentially free of native endogenous substances.
3. Human antithrombin III according to Claim 1 unaccompanied
10 by associated native glycosylation.
4. A DNA sequence comprising a sequence encoding human antithrombin III.
- 15 5. The DNA sequence according to Claim 4 including a sequence encoding the presequence peptide for human antithrombin III.
6. A replicable cloning vehicle containing the DNA sequence
20 according to Claims 4 or 5.
7. An expression vector comprising a DNA sequence encoding human antithrombin III operably linked to expression effecting DNA sequence and flanked by translational start
25 and stop signals.
8. A viable cell culture transfected with the expression vector of Claim 7.
- 30 9. A cell culture of Claim 8 capable of producing mature human antithrombin III.

-26-

10. Plasmids pATIII-E7 and pATIII-J4.
11. A cell culture transfected with each of the plasmids according to Claim 10.
- 5 12. A composition comprising a therapeutically effective amount of human antithrombin III according to Claim 1 in admixture with a pharmaceutically acceptable carrier.
- 10 13. A culture of cells altered so as to direct the genetic expression of human antithrombin III.
14. Human antithrombin III according to Claim 1 for use in
15 treating cardiovascular diseases or conditions or in pharmaceutical compositions useful for such treatment.
- 20 15. A process which comprises expressing a gene encoding human antithrombin III in a cell culture transfected with an expression vector operably harboring said gene.

1/5

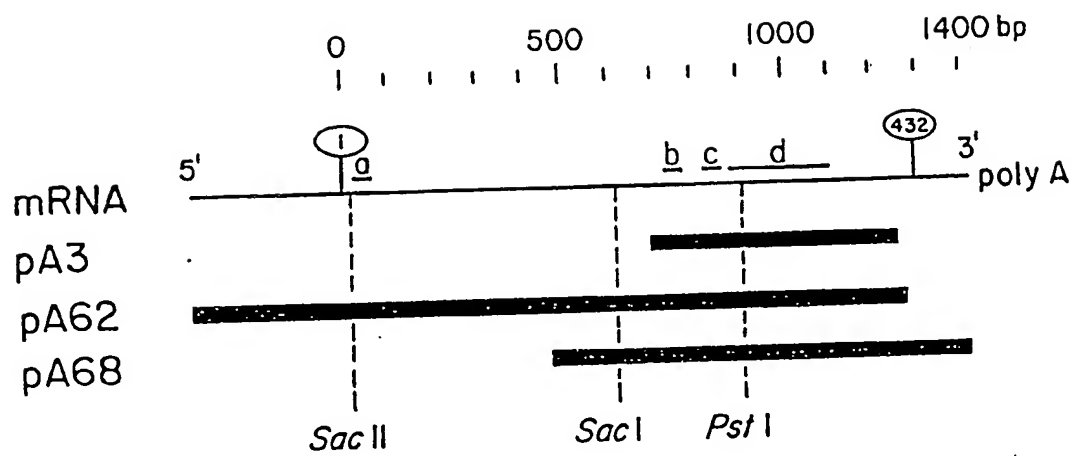


FIG 1

CACCAGCATCATCTCTCTCCAATTATCCAGCTACTCTGCCCCATGAGATAATAGTTTTCAGGCGGATTGCCICAGATCACACTATCTCCACTTGCCCCAGGCCCTGTGGAGATTAGCGGGC
 (Signal)
 Met Tyr Ser Asn Val Ile Gly Thr Val Thr Ser Gly Lys Arg Lys Val Tyr Leu Leu Ser Leu Leu Leu Ile Gly Phe Trp Asp Cys Val
 ATG TAT TCC AAT GTG ATA GGA ACT GTA ACC TCT GGA AAA AGG AAG GTT TAT CTT TTG TCC TTG CTG CTC ATT GGC TTC TGG GAC TGC GTG
 1 (Mature)
 Thr Cys His Gly Ser Pro Val Asp Ile Cys Thr Ala Lys Pro Arg Asp Ile Pro Met Asn Pro Met Cys Ile Tyr Arg Ser Pro Glu Lys
 ACC TGT CAC GGG AGC CCT GTG GAC ATC TGC ACA GCC AAG CCG CGG GAC ATT CCC ATG AAT CCC ATG TGC ATT TAC CGC TCC CCG GAG AAG
 50
 Lys Ala Thr Glu Asp Glu Gly Ser Glu Gln Lys Ile Pro Glu Ala Thr Asn Arg Arg Val Trp Glu Leu Ser Lys Ala Asn Ser Arg Phe
 AAG GCA ACT GAG GAT GAG GGC TCA GAA GAG AAG ATC CCG GAG GCC ACC AAC CGG CGT GTC TGG GAA CTG TCC AAG GCC AAT TCC CGC TTT
 Ala Thr Thr Phe Tyr Gln His Leu Ala Asp Ser Lys Asn Asp Asn Ile Phe Leu Ser Pro Leu Ser Ile Ser Thr Ala Phe Ala
 GCT ACC ACT TTC TAT CAG CAC CTG GCA GAT TCC AAG AAT GAC AAT GAT AAC ATT TTC CTG TCA CCC CTG AGT ATC TCC ACG GCT TTT GCT
 100
 Met Thr Lys Leu Gly Ala Cys Asn Asp Thr Leu Gln Gln Leu Met Glu Val Phe Lys Phe Asp Thr Ile Ser Glu Lys Thr Ser Asp Gln
 ATG ACC AAG CTG GGT GCC TGT AAT GAC ACC CTC CAG CAA CTG ATG GAG GTA TTT AAG TTT GAC ACC ATA TCT GAG AAA ACA TCT GAT CAG
 Ile His Phe Phe Phe Ala Lys Leu Asn Cys Arg Leu Tyr Arg Lys Ala Asn Lys Ser Ser Lys Leu Val Ser Ala Asn Arg Leu Phe Gly
 ATC CAC TTC TTC TTT GCC AAA CTG AAC TGC CGA CTC TAT CGA AAA GCC AAC AAA TCC TCC AAG TTA GTA TCA GCC AAT CGC CTT TTT GGA
 150
 Asp Lys Ser Leu Thr Phe Asn Glu Thr Tyr Gln Asp Ile Ser Glu Leu Val Tyr Gly Ala Lys Leu Gln Pro Leu Asp Phe Lys Glu Asn
 GAC AAA TCC CTT ACC TTC AAT GAG ACC TAC CAG GAC ATC AGT GAG TTG GTA TAT GGA GCC AAG CTC CAG CCC CTG GAC TTC AAG GAA AAT
 200
 Ala Glu Gln Ser Arg Ala Ala Ile Asn Lys Trp Val Ser Asn Lys Thr Glu Gly Arg Ile Thr Asp Val Ile Pro Ser Glu Ala Ile Asn
 GCA GAG CAA TCC AGA GCG GCC ATC AAC AAA TGG GTG TCC AAT AAG ACC GAA GGC CGA ATC ACC GAT GTC ATT CCC TCG GAA GCC ATC AAT
 Glu Leu Thr Val Leu Val Val Asn Thr Ile Tyr Phe Lys Gly Leu Trp Lys Ser Lys Phe Ser Pro Glu Asn Thr Arg Lys Glu Leu
 GAG CTC ACT GTT CTG GTG CTG GTT AAC ACC ATT TAC TTC AAG GGC CTG TGG AAG TCA AAG TTC AGC CCT GAG AAC ACA AGG AAG GAA GAA CTG

Fig 2 (first part)

250

Phe Tyr Lys Ala Asp Gly Glu Ser Cys Ser Ala Ser Met Met Tyr Gln Glu Gly Lys Phe Arg Tyr Arg Arg Val Ala Glu Gly Thr Gln
TTC TAC AAG GCT GAT GGA GAG TCG TGT TCA GCA TCT ATG ATG TAC CAG GAA GGC AAG TTC CGT TAT CCG CGC GTG GCT GAA GGC ACC CAG

Val Leu Glu Leu Pro Phe Lys Gly Asp Asp Ile Thr Met Val Leu Ile Leu Pro Lys Pro Glu Lys Ser Leu Ala Lys Val Glu Lys Glu
GTG CTT GAG TTG CCC TTC AAA GGT GAT GAC ATC ACC ATG ATG GTC CTC ATC TTG CCC AAG CCT GAG AAG AGC CTG GCC AAG GTG GAG AAG GAA

300

Leu Thr Pro Glu Val Leu Gln Glu Trp Leu Asp Glu Leu Glu Met Met Leu Val Val His Met Pro Arg Phe Arg Ile Glu Asp Gly
CTC ACC CCA GAG GTG CTG CAG GAG TGG CTG GAT GAA TTG GAG GAG ATG ATG CTG GTG GTT CAC ATG CCC CGC TTC CGC ATT GAG GAC GGC

350

Phe Ser Leu Lys Glu Gln Leu Glu Asp Met Gly Leu Val Asp Leu Phe Ser Pro Glu Lys Ser Lys Leu Pro Gly Ile Val Ala Glu Gly
TTC AGT TTG AAG GAG CAG CTG CAA GAC ATG GGC CTT GTC GAT CTG TTC AGC CCT GAA AAG TCC AAA CTC CCA GGT ATT GTT GCA GAA GGC

Arg Asp Asp Leu Tyr Val Ser Asp Ala Phe His Lys Ala Phe Leu Glu Val Asn Glu Gly Ser Glu Ala Ala Ser Thr Ala Val
CGA GAT GAC CTC TAT GTC TCA GAT GCA TTC CAT AAG GCA TTT CTT GAG GTA AAT GAA GAA GGC AGT GAA GCA GCT GCA AGT ACC GCT GTT

400

Val Ile Ala Gly Arg Ser Leu Asn Pro Asn Arg Val Thr Phe Lys Ala Asn Arg Pro Phe Leu Val Phe Ile Arg Glu Val Pro Leu Asn
GTG ATT GCT GGC CGT TCG CTA AAC CCC AAC AGG GTG ACT TTC AAG GCC AAC AGG CCC TTC CTG GTT TTT ATA AGA GAA GTT CCT CTG AAC

Thr Ile Ile Phe Met Gly Arg Val Ala Asn Pro Cys Val Lys End

ACT ATT ATC TTC ATG GGC AGA GTA GCC AAC CCT TGT GTT AAG TAA AAT GTT CTT ATT CTT TGC ACC TCT TCC TAT TTT TGG TTT GTG AAC

AGA AGT AAA AAT AAA TAC AAA CTA CTT CCA TCT CAC ATT Polya

Fig 2 (second part)

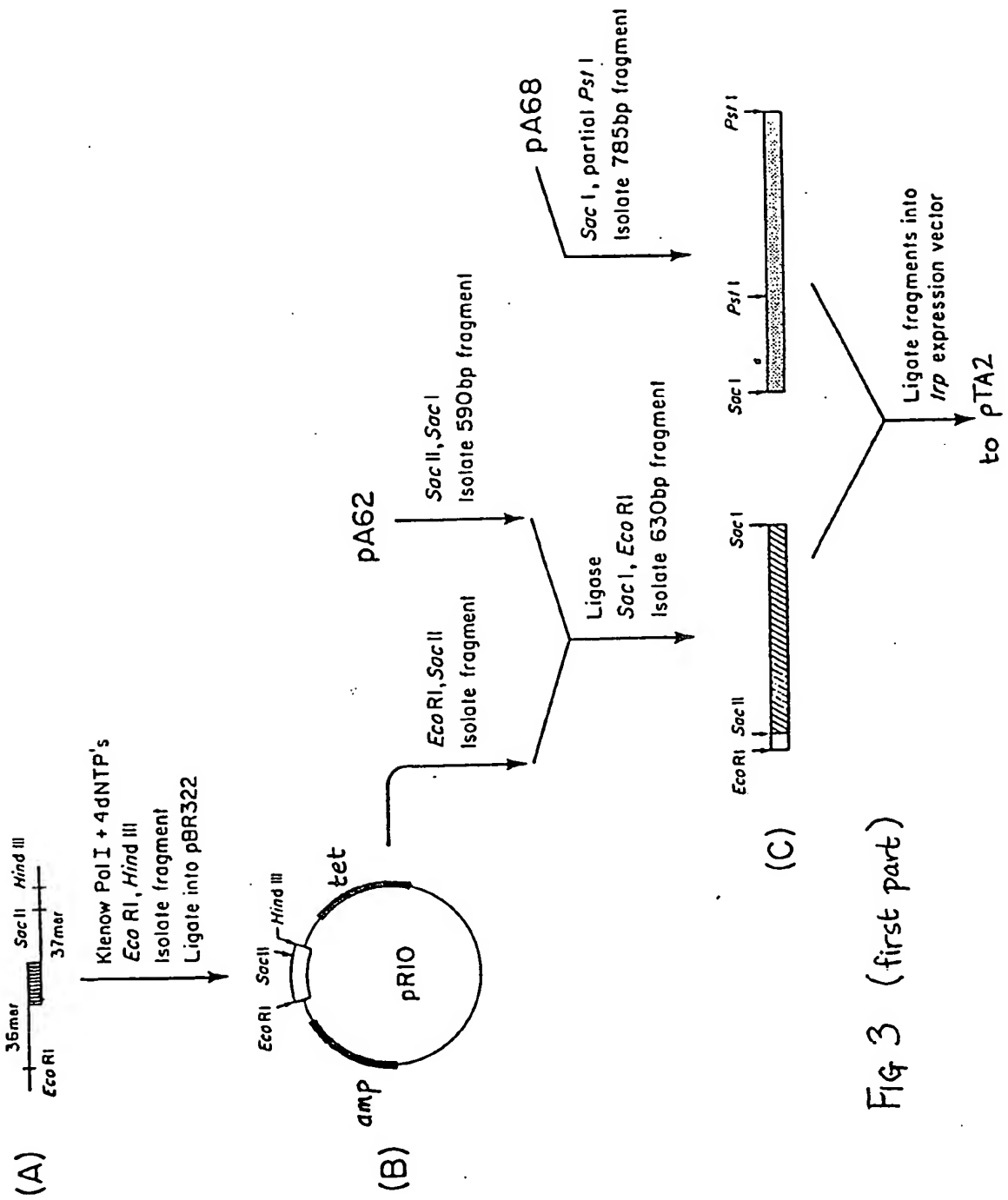


Fig 3 (first part)

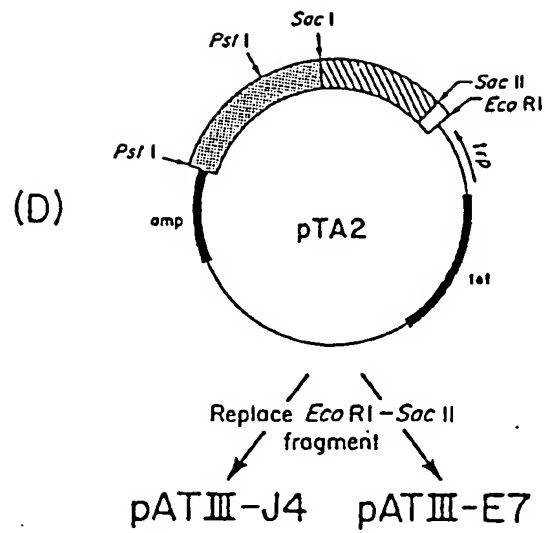


FIG 3 (second part)